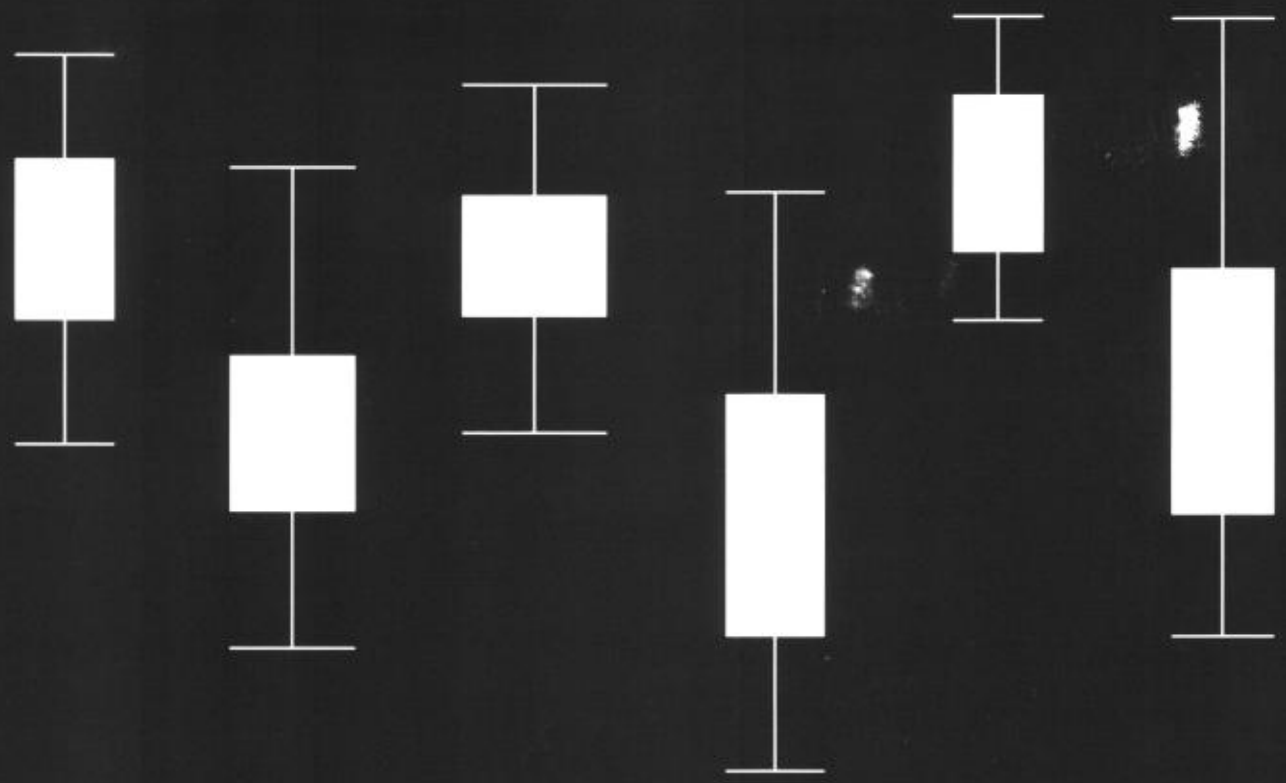
GRE® VALIDITY STUDY SERVICE



# VALIDITY OF THE GRE: 1988-89 SUMMARY REPORT

September 1990

Published for the Graduate Record Examinations Board
by Educational Testing Service

*Lisa M. Schneider*
*Jacqueline B. Briel*

# Validity of the GRE:

## 1988-89 Summary Report

Lisa M. Schneider
Jacqueline B. Briel

*Educational Testing Service is an Affirmative Action/Equal Opportunity Employer.*

## Foreword

The Validity Study Service was established in 1979 to help graduate schools and departments investigate the validity of the Graduate Record Examinations and other predictors of success in graduate school. It developed as an outgrowth of the Cooperative Validity Studies Project, a research effort funded by the GRE Board Research Committee from 1975 to 1978. The results of that project suggested that the usefulness of validity studies for individual departments can be increased by reporting them in the context of results for departments in the same field. The Validity Study Service is, therefore, designed to have a cooperative reporting function and an individual department reporting function.

This report is one in a series prepared by the GRE Validity Study Service. The Cooperative Validity Studies were designed and executed by Kenneth Wilson (1979). Previous summary reports were written by Neal Kingston, Samuel Livingston, Nancy Turner, and Nancy Burton. (Kingston and Livingston, 1981; Livingston and Turner, 1982; Burton and Turner, 1983).

Charlotte V. Kuh
GRE Program Director

Acknowledgements

The GRE Program would like to thank the following people at Educational Testing Service for providing comments on various drafts of this report: Robert Altman, Robin Durso, Neal Kingston, Charlotte Kuh, Charles Lewis, Samuel Livingston, Craig Mills, Kathleen O'Neill, Eldon Park, Nancy Petersen, Don Powers, and Warren Willingham.

We gratefully acknowledge the assistance of Ellen Callahan and Jennifer Thul of Educational Testing Service for their work on the Validity Study Service. Finally, we would like to express our sincere thanks to Manfred Steffen of Educational Testing Service for his valuable contributions in developing this report as well as for his technical advise regarding the operation of the Validity Study Service.

## Table of Contents

## EXECUTIVE SUMMARY

In recognition of its responsibility to assist institutions in the interpretation and use of GRE scores in the graduate admissions process, the GRE Board established the GRE Validity Study Service (VSS) in 1979. The VSS can be a useful tool in determining how information about applicants (predictor variables) can be used to make judgments about the probability of their success in graduate school (criterion variables).

The purpose of this report is twofold: 1) to describe the VSS and its participants and, 2) to summarize what we know about the validity of the GRE tests. The first part of this report summarizes the conclusions drawn from 606 validity studies that have been performed for graduate departments in the last five years and discusses the policy implications that might be useful in admissions decisions. The remainder of the report describes the service and the results from which these conclusions have been drawn.

This report is one in a series prepared by the GRE VSS. These reports provide a general summary of information that has been gathered through the service and represent only a small part of the validity information that is available. Comprehensive validity information based on these and other studies can be found in the forthcoming GRE Technical Handbook: Test Development, Score Interpretation, and Research for the Graduate Record Examinations Program.

### KEY FINDINGS

The main conclusions drawn from the analysis of the results can be summarized as follows:

1. GRE General Test scores tend to show moderate correlations with first-year averages. GRE Subject Test scores are even better predictors of first-year average than are the General Test scores. Of the eight GRE Subject Tests for which sufficient data are available, the GRE Subject Test scores are the best single predictors of first-year averages for half of the Subject Tests, while undergraduate grade point average (GPA) is the best single predictor for the other half.

2. The combination of GRE General Test scores and undergraduate GPA correlates more highly with first-year averages than either GRE scores alone or undergraduate GPA alone.

3. The GRE Subject Test scores add substantially to the predictive power of a set of predictors. The Subject Tests add more predictive power to undergraduate GPA than do GRE General Test scores.

4. Four-year cumulative undergraduate GPA, undergraduate GPA in major only, and undergraduate GPA in last two years have similar predictive power for first-year average. Undergraduate GPA in last two years predicts faculty rating better than either four-year cumulative undergraduate GPA or undergraduate GPA in major only. This issue will be investigated further as more data become available.

## POLICY IMPLICATIONS AND RECOMMENDATIONS

- A composite of GRE scores (verbal, quantitative, and analytical) and undergraduate GPA forms a useful basis for admissions decisions along with other information. However, simply summing scores to form a composite may misrepresent the relative importance of the predictors of graduate school success. The GRE Program recommends that if a department uses a weighted composite, the composite weights be obtained from a validity study for that individual department. Different departments may find, for example, that GRE verbal scores are more or less important determinants of success in their particular programs.

- The VSS provides expectancy tables for individual departments that participate in the service. These tables allow a department to assess the level of risk it is willing to take in the admissions process. For example, to increase diversity, a department may want to accept greater risks, as may be reflected in the tables, in admitting students from groups that have historically lacked access to their field, such as women and minorities.

- Special consideration is required in admitting students who communicate best in a language other than English. For example, while the GRE verbal score might be useful in comparing students within this group, it is not always useful for comparing such students with others who do communicate better in English. Further, research on the Test of English as a Foreign Language (TOEFL) has shown that foreign students must demonstrate some minimal level of English proficiency before their scores on verbal aptitude tests begin to become meaningful (Angelis, Swinton, & Cowell, 1979; Alderman, 1981).

## GRE VALIDITY STUDY SERVICE SUMMARY

### OVERVIEW

The GRE VSS is a free service whereby graduate departments can submit data to evaluate the usefulness of several variables, such as GRE scores and undergraduate GPA, for predicting success in graduate school. This information can be used to determine which variables should be incorporated into the admissions process and how they might be used. The VSS, therefore, provides institutions with a measure of predictive validity. There are several other types of validity that are not addressed in this report. A summary of the different types of validity can be found in Educational Measurement (Thorndike, 1971).

Departments with a minimum of five first-time, degree-seeking students are eligible to participate in the VSS. Participating departments submit data for students entering graduate school within a specified period of time. Only students within a department, taking course work of the same type and level, are included so that the criterion will be directly comparable for all students in the study.

Because the GRE tests were originally developed for English-speaking students and may function differently for those students whose best language is not English, analyses are provided separately for two groups: students who communicate better in English and those who communicate better in some other language. Departments may supply data for one or both groups; however, each group must contain a minimum of five students. When departments do supply data for both groups, each group is considered separately in all analyses.

The GRE VSS uses a statistical technique called empirical Bayes regression[1]. Because the empirical Bayes method makes use of data from all participating departments to perform validity studies for individual departments, the VSS requires that data be available in the VSS database for a minimum of 100 students across ten departments for each predictor or criterion variable to be reported.

#### Predictors of Academic Success

The standard predictors used in the VSS are GRE General Test scores (verbal, quantitative, and analytical), four-year cumulative undergraduate GPA, and, at the department's request, GRE Subject Test scores. Because undergraduate grades are not directly comparable for foreign institutions, undergraduate GPA is not analyzed for the group of students who do not communicate better in English (under the assumption that many of these students may have attended non-U.S. institutions). Beginning with the 1987-88 cycle of validity studies, departments were allowed to choose one of two optional predictors to include in their departmental study. These two optional variables are undergraduate GPA in major only and undergraduate GPA for last two years. Of the 329 departments that participated in the VSS between 1987-89, 53 requested analyses for undergraduate GPA in major only, and 82 chose undergraduate GPA for last two years.

#### Criteria of Academic Success

The standard criterion of academic success for the VSS is first-year average in graduate school. Beginning with the 1987-88 cycle, a faculty rating scale was predefined as an optional criterion variable. Of the 329 departments that participated in the VSS between 1987-89, a total of 89 departments chose to provide data for the optional criterion, faculty rating.

Although first-year average has some weaknesses as a criterion, it is useful for several reasons. First, a student's first-year average represents the combined judgment of his or her instructors. Although individual grading standards may vary among instructors and courses may vary in difficulty, the differences tend to average out when grades are accumulated. Second, grades represent an official evaluation of a student's performance and are available in nearly every graduate department. Finally, first-year average is a measure that is readily available, whereas other criteria may require a considerable amount of time to obtain information that can be used in a validity study.

---

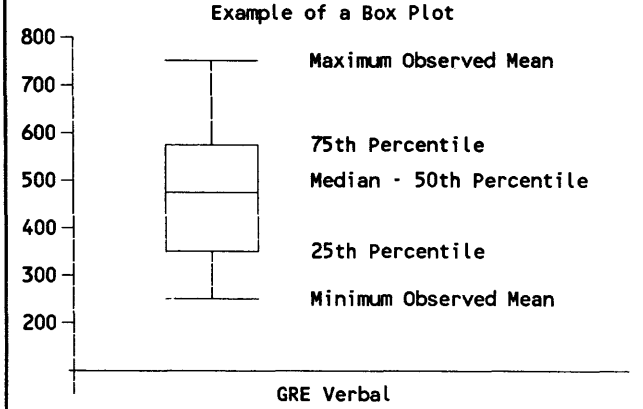
[1] Technical details regarding this method can be found in Braun and Jones, (1985). The method is also described in Appendix A of this report.

Using faculty rating as a measure of success in graduate school has both strengths and weaknesses. The four-point rating scale allows faculty to give students a more global rating, one that includes other elements in addition to grades. Instructors can also differentiate better among students with varying ability levels since they are more likely to use the entire scale. One disadvantage of such a rating system, however, is the possible occurrence of a halo effect or other rating effects, which could occur when a rater allows extraneous impressions to influence the rating. For example, if we like a person, we might tend to rate that person higher on academic performance measures, although personality characteristics are not necessarily related to academic performance. The VSS recommends that departments average all faculty ratings to create one composite rating for a student.

Average Departmental Performance

Figures B.1 through B.11 in Appendix B are graphical representations of departmental means. The data are displayed in the form of box and whiskers plots. Box plots are intended to visually summarize a score distribution by plotting five selected values. By way of example, consider the fictitious distribution of GRE Verbal means in Figure 1. The vertical boundaries of the "box" are drawn at the 25th and 75th percentiles (350 and 575, respectively). As such, 50% of the means fall within this box. The median (50th percentile) is depicted by the horizontal line within the box. In this example, the median is 475. At this point we have described not only the middle 50% of the means but also the shape of the distribution. This distribution of means is slightly asymmetrical as indicated by the larger distance between the median and the 25th percentile than between the median and the 75th percentile. The "whiskers" of the plot are drawn to extend from the 75th percentile to the maximum mean (760) and from the 25th percentile to the minimum mean (260). The minimum and maximum means are represented by the horizontal lines. The whiskers further assist in determining the shape of the distribution of departmental means.

Figure 1
Example of a Box Plot




The individual reports received by departments participating in the VSS also contain box plots. Those box plots, however, are based on student scores for that particular department, whereas the box plots in this report are based on means for each department. Because the values presented in these two reports are not the same (individual scores vs. departmental means), the plots provided in the individual departmental reports and those in this report should not be compared.

## RESULTS

The GRE VSS has collected and analyzed data for many individual departments since the establishment of the service in 1979. As a result, the VSS database is a valuable source of validity information. In an effort to provide a global picture of the validity of the GRE tests, this study is intended to summarize the validity information currently available.

Individual departments provide data, such as undergraduate GPA, first-year average, and descriptive information for first-time, degree-seeking students. In addition to these data, departments also provide GRE scores to aid in the location of the students in the GRE computer files. However, the score data actually used in the analyses are taken directly from the computer files. The GRE computer files include data for the past five testing years. Therefore, the analyses in this report are based on data for students who have taken the GRE tests between October 1983 and September 1988.

Analyses are provided by language proficiency group for all departments combined and for departments in major field areas where the database sample size requirements are fulfilled (a minimum of 100 students across at least ten departments). Table 1 provides a detailed classification of the types of departments participating in the VSS. A further breakdown of the participating departments by sex, ethnicity, age, and department size is provided in Appendix C.

### Analysis Questions

The purpose of the GRE VSS is to provide information concerning the usefulness of GRE scores and undergraduate GPA in the admissions process. Listed below are the specific questions addressed in this report and a brief answer for each. Following each question is a listing of the tables that provide supporting data.

1) **How well do GRE General Test scores predict success in graduate school?**
**(Tables 2, 3, 4, and 5)**
GRE General Test scores tend to show moderate correlations with first-year average.

2) **How does the predictive power of GRE General Test scores compare with that of undergraduate GPA in terms of graduate school success?**
**(Tables 2 and 4)**
The combination of GRE General Test scores and undergraduate GPA predicts first-year average more effectively than either GRE scores alone or undergraduate GPA alone.

3) **What contribution do GRE Subject Test scores make in predicting graduate school success?**
**(Tables 7 and 8)**
GRE Subject Test scores are better predictors than are the combined GRE General Test scores and in half of the cases, better predictors than undergraduate GPA. GRE Subject Test scores also add substantially to the predictive power of a set of predictors.

4) **Are there differences in the predictive power of GRE General Test scores for students who communicate better in English and those who do not?**
**(Tables 2 and 4 compared to Tables 3 and 5)**
Correlations indicate differences in the relative effectiveness of the predictors for the two subgroups.

5) **What factors, in the study design and in the graduate admissions process in general, affect the evaluation of the contribution of GRE scores to prediction?**
**(Table 9)**
Restriction of range in the variables and compensatory selection procedures might limit the validity estimates.

# Table 1
## Participants in the GRE Validity Study Service Classified by Department Type

Validity Sample : Students Who Communicate Better in English

| Type of Department | Number of Depts. | Number of Students |
|---|---|---|
| All Departments | 606 | 9,200 |
| Natural Sciences | 229 | 3,069 |
| Agriculture | 19 | 184 |
| Biological Sciences | 62 | 650 |
| Chemistry | 22 | 340 |
| Computer & Information Sci. | 21 | 252 |
| Earth, Atmos., & Marine Sci. | 28 | 392 |
| Health & Medical Sciences | 53 | 1,017 |
| Mathematical Sciences | 14 | 101 |
| Physics & Astronomy | 9 | 119 |
| Other | 1 | 14 |
| Engineering | 34 | 442 |
| Chemical | 8 | 107 |
| Civil | 3 | 58 |
| Electrical & Electronics | 7 | 71 |
| Industrial | 2 | 26 |
| Mechanical | 6 | 90 |
| Other | 8 | 90 |
| Social Sciences | 203 | 3,365 |
| Anthropology & Archaeology | 5 | 56 |
| Communications | 16 | 219 |
| Economics | 26 | 261 |
| Home Economics | 6 | 55 |
| Library & Archival Sciences | 8 | 263 |
| Political Science | 20 | 338 |
| Psychology | 70 | 1,233 |
| Public Administration | 12 | 283 |
| Social Work | 3 | 85 |
| Sociology | 14 | 143 |
| Other | 23 | 429 |
| Humanities & Arts | 74 | 1,266 |
| Architect. & Environ. Design | 6 | 171 |
| Arts: History,Theory, & Crit. | 3 | 40 |
| Arts: Performance & Studio | 9 | 131 |
| English Language & Lit. | 18 | 269 |
| Foreign Languages & Lit. | 4 | 39 |
| History | 17 | 165 |
| Philosophy | 6 | 72 |
| Religion & Theology | 7 | 331 |
| Other | 4 | 48 |
| Education | 58 | 960 |
| Administration | 7 | 95 |
| Curriculum & Instruction | 8 | 180 |
| Elementary | 5 | 70 |
| Evaluation & Research | 4 | 54 |
| Secondary | 4 | 35 |
| Special | 3 | 56 |
| Student Couns. & Personnel | 4 | 51 |
| Other | 23 | 419 |
| Business | 8 | 98 |
| Business Admin. & Management | 6 | 82 |
| Other | 2 | 16 |

Validity Sample : Students Who Do Not Communicate Better in English

| Type of Department | Number of Depts. | Number of Students |
|---|---|---|
| All Departments | 100 | 960 |
| Natural Sciences | 40 | 361 |
| Agriculture | 2 | 14 |
| Biological Sciences | 6 | 56 |
| Chemistry | 8 | 92 |
| Computer & Information Sci. | 16 | 142 |
| Earth, Atmos., & Marine Sci. | 2 | 10 |
| Health & Medical Sciences | 2 | 15 |
| Mathematical Sciences | 2 | 18 |
| Physics & Astronomy | 2 | 14 |
| Engineering | 31 | 329 |
| Chemical | 3 | 34 |
| Civil | 4 | 66 |
| Electrical & Electronics | 8 | 55 |
| Industrial | 2 | 35 |
| Mechanical | 5 | 73 |
| Other | 9 | 66 |
| Social Sciences | 26 | 249 |
| Economics | 15 | 165 |
| Library & Archival Sciences | 1 | 6 |
| Political Science | 6 | 56 |
| Sociology | 2 | 12 |
| Other | 2 | 10 |
| Humanities & Arts | 3 | 21 |
| Architect. & Environ. Design | 1 | 6 |
| English Language & Lit. | 1 | 7 |
| Foreign Languages & Lit. | 1 | 8 |

Correlational Analyses

A statistic called the correlation coefficient summarizes the relationship between examination scores and first-year average (i.e., the extent to which students with higher scores tend to have higher first-year averages). Correlations obtained in the VSS tend to be positive and can range from zero to one. The closer the correlation is to one, the better the relationship; however, perfect correlations are rarely (or never) obtained in human measurement. Because a number of factors described later in this report act to limit the magnitude of correlations in predictive validity studies, correlations in the VSS tend to be in the .2 to .5 range.

The correlations reported in the tables are weighted correlations that take into account the differences in the number of students between departments. For example, a correlation in a department with 10 students contributes twice as much to the weighted correlation as a correlation based on 5 students. Correlations are provided for single predictors and various combinations of multiple predictors. Correlations tend to increase as additional variables are included in the analysis. The predictors are combined by creating a weighted sum through the multiplication of each predictor by specific weights to produce a composite score for each applicant (see Appendix A for a description of the method used to determine the weights). Any two or more predictors can be combined through these methods and the composite then can be correlated with first-year average.

Tables 2 through 5 provide correlations for the GRE General Test scores and undergraduate GPA separately and for various combinations of these predictors, with first-year average and faculty rating. One particularly interesting finding is that for all departments combined in the subgroup that does not communicate better in English, the predictors tend to predict faculty rating considerably better than they predict first-year average. This finding may be considered tentative, however, since sample size differences are large.

**Table 2**

**Size-Adjusted Average Correlations of Predictors with First-Year Average**

Validity Sample : Students Who Communicate Better in English

| Type of Department | Number of Depts. | Number of Students | Predictors* | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | V | Q | A | U | VQA | VQAU |
| All Departments | 606 | 9,200 | .29 | .28 | .26 | .34 | .33 | .43 |
| Natural Sciences | 229 | 3,069 | .26 | .24 | .23 | .33 | .29 | .41 |
| Engineering | 34 | 442 | .22 | .18 | .20 | .35 | .25 | .40 |
| Social Sciences | 203 | 3,365 | .31 | .32 | .28 | .33 | .36 | .45 |
| Humanities & Arts | 74 | 1,266 | .30 | .31 | .28 | .39 | .34 | .47 |
| Education | 58 | 960 | .29 | .29 | .26 | .29 | .33 | .42 |

* V = GRE Verbal, Q = GRE Quantitative, A = GRE Analytical, U = Undergraduate GPA

**Table 3**

**Size-Adjusted Average Correlations of Predictors with First-Year Average**

Validity Sample : Students Who Do Not Communicate Better in English

| Type of Department | Number of Depts. | Number of Students | Predictors* | | | |
|---|---|---|---|---|---|---|
| | | | V | Q | A | VQA |
| All Departments | 100 | 960 | .23 | .31 | .25 | .39 |
| Natural Sciences | 40 | 361 | .23 | .30 | .25 | .40 |
| Engineering | 31 | 329 | .22 | .29 | .25 | .36 |
| Social Sciences | 26 | 249 | .23 | .34 | .25 | .41 |

* V = GRE Verbal, Q = GRE Quantitative, A = GRE Analytical

**Table 4**

**Size-Adjusted Average Correlations of Predictors with Faculty Rating**

Validity Sample : Students Who Communicate Better in English

| Type of Department | Number of Depts. | Number of Students | Predictors* | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | V | Q | A | U | VQA | VQAU |
| All Departments | 89 | 891 | .25 | .25 | .21 | .31 | .37 | .44 |
| Natural Sciences | 42 | 401 | .25 | .23 | .22 | .34 | .36 | .46 |
| Social Sciences | 30 | 320 | .26 | .25 | .21 | .30 | .36 | .42 |
| Humanities & Arts | 10 | 107 | .26 | .31 | .22 | .30 | .47 | .47 |

* V = GRE Verbal, Q = GRE Quantitative, A = GRE Analytical, U = Undergraduate GPA

**Table 5**

**Size-Adjusted Average Correlations of Predictors with Faculty Rating**

Validity Sample : Students Who Do Not Communicate Better in English

| Type of Department | Number of Depts. | Number of Students | Predictors* | | | |
|---|---|---|---|---|---|---|
| | | | V | Q | A | VQA |
| All Departments | 15 | 143 | .34 | .40 | .38 | .56 |

* V = GRE Verbal, Q = GRE Quantitative, A = GRE Analytical

Comparison of the correlations for the two language proficiency subgroups indicates differences in the relative effectiveness of the predictors. The verbal score predicts first-year average slightly better for most fields for the subgroup that communicates better in English, while the quantitative score has more predictive power for the subgroup that does not communicate better in English. It should be noted that the combination of the three GRE General Test scores (undergraduate GPA is not analyzed for the subgroup that does not communicate better in English) predicts both first-year average and faculty rating better for the subgroup that does not communicate better in English than for the subgroup that communicates better in English.

Table 6 provides correlations of the optional predictors with first-year average and faculty rating. The two optional predictors tend to predict first-year average about equally well, while undergraduate GPA in the last two years tends to predict faculty rating better than does undergraduate GPA in major only.

**Table 6**

**Size-Adjusted Average Correlations of Optional Predictors with Criteria**

Validity Sample : Students Who Communicate Better in English

| Optional Predictor | Criterion | Number of Depts. | Number of Students | Correlation |
|---|---|---|---|---|
| UGPA - In Major Only | | | | |
| | First-Year Average | 53 | 652 | .35 |
| | Faculty Rating | 29 | 290 | .28 |
| UGPA - Last Two Years | | | | |
| | First-Year Average | 82 | 863 | .32 |
| | Faculty Rating | 31 | 266 | .40 |

Correlations of the Subject Tests with the criteria are provided in Tables 7 and 8. It should be noted that the GRE Subject Tests (S) substantially add to the predictive power of a set of predictors. A comparison across all Subject Tests shows that the addition of S to a combination of predictors produces considerably larger correlations than the same combinations without S (see specifically columns VQASU and VQAU). For example, consider the Engineering Test in Table 7. The correlation for the combination of GRE General Test scores and undergraduate GPA is .41 while the addition of the Subject Test to this set of predictors yields a correlation of .60. Furthermore, the correlations in Table 7 also suggest that the Subject Tests generally make a larger contribution to the prediction of success in graduate school than do the General Test scores. This can be observed by comparing columns S and VQA, as well as columns SU and VQAU.

Table 7

**Size-Adjusted Average Correlations of Predictors with First-Year Average for Eight Subject Tests**

Validity Sample : Students Who Communicate Better in English

| GRE Subject Test | Number of Depts. | Number of Students | Predictors* | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | V | Q | A | S | U | SU | VQA | VQAS | VQAU | VQASU |
| Biology | 41 | 354 | .24 | .23 | .22 | .27 | .32 | .45 | .28 | .32 | .39 | .48 |
| Chemistry | 23 | 297 | .25 | .23 | .23 | .48 | .36 | .52 | .28 | .53 | .41 | .57 |
| Economics | 15 | 129 | .23 | .26 | .27 | .42 | .34 | .52 | .29 | .49 | .41 | .57 |
| Engineering | 12 | 116 | .19 | .14 | .19 | .33 | .39 | .56 | .21 | .34 | .41 | .60 |
| Geology | 15 | 246 | .15 | .18 | .20 | .21 | .28 | .31 | .20 | .34 | .32 | .41 |
| Literature in English | 15 | 174 | .26 | .32 | .27 | .37 | .40 | .53 | .32 | .45 | .44 | .57 |
| Physics** | 10 | 119 | .18 | .14 | .20 | .43 | .37 | .53 | .21 | .51 | .39 | .57 |
| Psychology | 58 | 940 | .26 | .25 | .24 | .36 | .34 | .44 | .30 | .42 | .40 | .48 |

* V = GRE Verbal, Q = GRE Quantitative, A = GRE Analytical, S = GRE Subject Test, U = Undergraduate GPA

** Correlations based on data from previous analyses.

Table 8

**Size-Adjusted Average Correlations of Predictors with Faculty Rating for the Psychology Subject Test**

Validity Sample : Students Who Communicate Better in English

| GRE Subject Test | Number of Depts. | Number of Students | Predictors* | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | V | Q | A | S | U | SU | VQA | VQAS | VQAU | VQASU |
| Psychology | 13 | 152 | .23 | .23 | .22 | .26 | .26 | .38 | .32 | .49 | .36 | .54 |

* V = GRE Verbal, Q = GRE Quantitative, A = GRE Analytical, S = GRE Subject Test, U = Undergraduate GPA

## ASSESSING THE LIMITATIONS OF PREDICTIVE VALIDITY STUDIES

Some limitations for the correlations between the predictors and the criteria, as noted in this report, must be considered. Please note, however, that most often these limitations serve to reduce or limit the correlations rather than to inflate them artificially.

- All predictors and criteria have some degree of unreliability. The reliability of these variables places an upper limit on obtainable validity coefficients.

- Students who attended graduate school and are thus included in the analysis do not represent either all those who applied or all those who were admitted. Therefore, the range of attendees' abilities is narrower than for the group from which selections are made. This "restriction of range" tends to lower correlations from what they would be in the applicant group.

- The analyses include only those students for whom complete data are available. This restriction may affect overall correlations in ways that are not usually ascertainable.

- To the extent that admissions decisions generally provide for compensatory selection of students, high scores on one predictor compensate for low scores on another. Although this is a common practice, it may reduce observed correlations.

- Poorer students often fail to complete their first year of graduate school, thus further restricting the range of first-year average. At the other extreme, in some departments more than half the students have 4.0 averages. Restrictions in criterion range, like those in predictor range, tend to further decrease the magnitude of correlations.

- First-year average is based on different sets of courses for different students, making it less comparable across students and generally reducing correlations.

We can examine further the effect of restriction of range and compensatory selection practices. An admission test score or other predictor that would correlate strongly with first-year average in the full group of applicants will tend to correlate less strongly with first-year average in a group of students with a narrower range of abilities. Of course, many students with low admission scores are not admitted. At the other extreme, those with higher scores who are accepted may not attend, limiting the range of predictors at the other end of the spectrum. Finally, some students with low scores on the GRE tests are admitted because of compensatory performance on other measures, such as undergraduate GPA. However, these students are not representative of all who score low on the test, and this selectivity will cause the correlation between test score and first-year average to be smaller than it would be if the students admitted were a representative sample of those who apply.

It would be desirable to estimate the correlations that might be obtained had no selection taken place. However, this cannot be accomplished without information for applicants who were rejected, did not apply, or did not enroll. In order to evaluate the effects of restriction of range and compensatory selection with the available data, it was necessary to identify departments that possibly exhibited these influences. Evidence of restriction of range was determined by a small standard deviation in the predictor variables (V, Q, or A). Compensatory selection was determined by a low (or negative) correlation between the predictor variables and undergraduate GPA.

Table 9 presents the results of this analysis. Departments classified as having "Severe" selection effects were those that, relative to all departments, had smaller standard deviations for the predictor and lower correlations between the predictor and undergraduate GPA. Departments classified as having "Slight" selection effects were those that had larger standard deviations and larger correlations relative to all departments. As a result, not every department is represented in this table. The correlations in the table are based on a combination of GRE scores and undergraduate GPA (VQAU) for only the departments that were determined to have the desired characteristics. Two major observations can be made. First, correlations in the "Severe" columns are substantially lower than those in the "Slight" columns. These data provide evidence that selection effects, as defined in this analysis, reduce observed validity correlations. Second, the overall average correlations (repeated from the VQAU column in Table 2) are smaller than those in the "Slight" column. This is evidence that the validity coefficients presented in this report are underestimates of the predictive potential of the predictor variables, and in fact the true validity of the predictors (i.e., the validity for the applicant pool) is closer to .50 than to .40.

**Table 9***

**Analysis of Selection Effects**

Validity Sample : Students Who Communicate Better in English

| | | Relative Amount of Selection Effects by Predictor | | | | | |
|---|---|---|---|---|---|---|---|
| | | Verbal | | Quantitative | | Analytical | |
| Type of Department | Overall Average Correlation | Severe | Slight | Severe | Slight | Severe | Slight |
| All Departments | .43 | .34 | .55 | .37 | .49 | .35 | .49 |
| Natural Sciences | .41 | .36 | .52 | .38 | .49 | .34 | .49 |
| Social Sciences | .45 | .34 | .56 | .37 | .51 | .33 | .53 |
| Humanities & Arts | .47 | ** | .56 | ** | .46 | .40 | .50 |
| Education | .42 | ** | .44 | ** | ** | ** | .41 |

* See text for description of this table.

** Only correlations based on 100 or more students are presented.

Low correlations are not necessarily indicative of weak relationships with performance. A particular variable may appear unimportant when in fact it may be the most important variable in terms of admissions. For example, consider the conventional wisdom that quantitative skills are important determiners of success in engineering. However, all applicants for an engineering department tend to have highly developed quantitative skills. As a consequence, the effectiveness of any quantitative measure is limited by the nature of the candidate pool itself. Furthermore, this phenomenon is likely to enhance the relative effectiveness of, for example, verbal measures. Thus the correlation associated with the quantitative measure may be low relative to the correlation associated with the verbal measure, and hence appear to be a weaker predictor of success.

## SUMMARY

The data collected through the VSS and analyzed in this report indicate that the GRE tests can be a useful part of the admissions process. The combination of GRE General Test scores, Subject Test scores, and undergraduate GPA has predictive power in spite of factors that tend to mask the true value of the tests, such as restriction of range and compensatory selection practices. Although the single predictors provide interesting information, data from the combined predictors should be used in practice.

These results do not suggest, however, that GRE test scores and undergraduate GPA should be used as a formula for admissions. Rather, it implies that GRE scores can be a useful part of the admissions process, even when other factors are being considered. This can be especially important when a department is attempting to diversify its student body by (1) providing some information about the level of risk associated with their efforts and (2) helping to identify students who should be targeted early for support to increase their chances of success.

The main conclusions of this report can be summarized as follows:

- The use of GRE General Test and Subject Test scores, in addition to undergraduate GPA, provides more effective prediction than undergraduate GPA alone.

- The combination of GRE General Test scores and undergraduate GPA predicts first-year average more effectively than either GRE scores alone or undergraduate GPA alone.

- The GRE Subject Test scores add substantially to the predictive power of a set of predictors.

- GRE General Test scores tend to show moderate correlations with first-year graduate grade-point average. GRE Subject Test scores are better predictors than the combined GRE General Test scores and in some cases, better predictors than undergraduate GPA.

- Four-year cumulative undergraduate GPA, undergraduate GPA in major only, and undergraduate GPA in last two years appear to have similar predictive power for first-year average.

- Selection procedures for admitting students to graduate programs tend to make the correlations of GRE scores and undergraduate GPA with first-year average lower than they would be with an unselected group of applicants.

For further information on other GRE validity research or for information on how to participate in the GRE VSS, please write or call: Project Director, GRE Validity Study Service, Educational Testing Service, Mail Stop 22-U, Rosedale Road, Princeton, New Jersey 08541; (609) 243-8286.

## References

Alderman, D. L. (1981). Language proficiency as a moderator variable in testing academic aptitude (TOEFL Research Report No. 10, ETS Research Report RR 81-41). Princeton, NJ: Educational Testing Service.

Angelis, P. J., Swinton, S. S. & Cowell, W. R. (1979). The performance of non-native speakers of English on TOEFL and verbal aptitude tests (TOEFL Research Report No. 3, ETS Research Report RR 79-7). Princeton, NJ: Educational Testing Service.

Braun, H. I., & Jones, D. H. (1985). Use of empirical Bayes methods in the study of the validity of academic predictors of graduate school performance (GRE Research Report No. 79-13P). Princeton, NJ: Educational Testing Service.

Burton, N. W., & Turner, N. J. (1983). Effectiveness of the Graduate Record Examinations for predicting first-year grades. 1981-82 Summary Report of the Graduate Record Examinations Validity Study Service. Princeton, NJ: Educational Testing Service.

Educational Testing Service. (in preparation). GRE technical handbook: Test development, score interpretation, and research for the Graduate Record Examinations Program (2nd ed.). Princeton, NJ: Author.

Kingston, N. M., & Livingston, S. A. (1981). Effectiveness of the Graduate Record Examinations for predicting first-year grades. 1979-80 Summary Report of the Graduate Record Examinations Validity Study Service. Princeton, NJ: Educational Testing Service.

Livingston, S. A., & Turner, N. J. (1982). Effectiveness of the Graduate Record Examinations for predicting first-year grades. 1980-81 Summary Report of the Graduate Record Examinations Validity Study Service. Princeton, NJ: Educational Testing Service.

Thorndike, R. L. (Ed.). (1971). Educational measurement (2nd ed.). Washington, DC: American Council on Education.

Wilson, K. M. (1979). The validation of GRE scores as predictors of first-year performance in graduate study: Report of the GRE cooperative validity studies project. (GRE Research Report No. 75-8R). Princeton, NJ: Educational Testing Service.

# Appendix A

## Description of the Empirical Bayes Method

The validity of a predictor is usually defined as the correlation between the predictor and the criterion it is meant to predict. Ordinarily, the validity is estimated by collecting data from a number of individuals with scores on both the predictor and the criterion and then applying standard statistical estimation techniques, such as least squares regression.

When performance in graduate school is the criterion, the estimation of validity is somewhat more complicated. First, the number of individuals in a given department is usually rather small. In addition, the interest is usually in estimating the joint validity of several predictors (e.g., V, Q, A, and U). When sample sizes are small and multiple predictors are involved, the estimates of the relationships among the criterion and the predictors are not well determined. Therefore, there is some uncertainty remaining about the true values of the coefficients and, correspondingly, about the true values of the validities.

The estimation of validity is further complicated by the fact that the students enrolled in the department are a small subset of the original set of applicants. Moreover, they have been selected in part on the basis of the predictors whose validity is to be assessed. Consequently, the range of variation in the predictors for the attending students is usually less than that of the entire group of applicants. This restriction of range also inflates the variance of the estimated validity coefficients.

One way to lessen the effects of these problems is to apply Bayesian statistical procedures. The particular version employed in the GRE VSS is called the empirical Bayes method. The essential idea is to treat the problem of estimating the validity of a particular set of predictors in a large number of departments as a set of related tasks. By appropriately structuring this view, it is possible to use the information gathered in all the other departments to aid in the estimation of the validity in any particular department.

How does empirical Bayes (EB) work? It begins by calculating the conventional least squares estimate of the regression plane in each department. The regression plane summarizes the association between first-year average (FYA) on the one hand and the various predictors on the other hand. The location and orientation of the regression plane are characterized by the regression coefficients.

The next step is to examine the relationship, if any, between the set of coefficients for a department and other departmental characteristics, such as the academic preparation of the students in the department (as measured, for example, by mean test scores) or the amount of variability between students on the various predictors. It is important to recognize that this phase of the analysis is not carried out at the same level as the first phase. Here the department is the basic unit of interest, whereas before it was the student within the department.

The relationship between the regression coefficients and the departmental characteristics is captured by a set of numbers that are referred to as the "gamma matrix." The gamma matrix may indicate, for example, that departments with higher-scoring students generally have a stronger association between FYA and the predictor scores.

The final step is to combine the information from the first two steps to produce a better estimate of each department's regression plane. Using the gamma matrix, regression coefficients typical of a department with a particular set of characteristics are calculated. This represents the best prediction when data from all departments are pooled. The EB estimate is a compromise between the classical estimate based on the department's data alone and the estimate derived from the gamma matrix. In effect, the classical estimate is shrunk toward the gamma matrix estimate, with the amount of shrinkage determined by the relative precision of the two estimates.

In general, for a small department the classical estimate is not well determined (i.e., the estimate has low precision). The EB estimate will then be very much like the prediction obtained from the gamma matrix, which is well determined because it is based on pooled data from all departments. On the other hand, the classical estimate for a large department is relatively well determined and, therefore, the EB estimate will not move too far in the direction of the gamma matrix prediction. The amount of shrinkage observed is an indication of how much of the variability between departments in the classical estimates is probably due to "noisy data" and how much is "real."

Although this approach may seem slightly counterintuitive at first, there are good methodological and substantive reasons for pursuing it. Although departments are unique and have some idiosyncrasies, there are many commonalities as well. Moreover, many of the observed idiosyncrasies are as likely to be due to sampling fluctuations as to real differences between that department and others similar to it. The empirical Bayes paradigm provides one way of deciding to what extent unusual results should be treated as genuine.

The obvious benefit of employing empirical Bayes is that estimates of the relationship between the criterion and the predictors (and therefore the joint validity as well) are more stable and have superior statistical properties than do estimates based on only one department's data. That superiority has been observed in a large number of investigations and agrees with the predictions of statistical theory.

One result is that the GRE VSS is now able to accommodate departments with fewer students and yet produce more detailed descriptions of the predictor-criterion relationships that exist for the students in the department. By combining information available for the individual department with the data from other departments, greater accuracy of estimation is achieved that in the classical approach would be possible only with much larger samples.

# Appendix B

## Box Plots Of Departmental Means For Individual Variables

## Figures B.1 - B.5

## Validity Sample : Students Who Communicate Better in English

*Figure B.1*

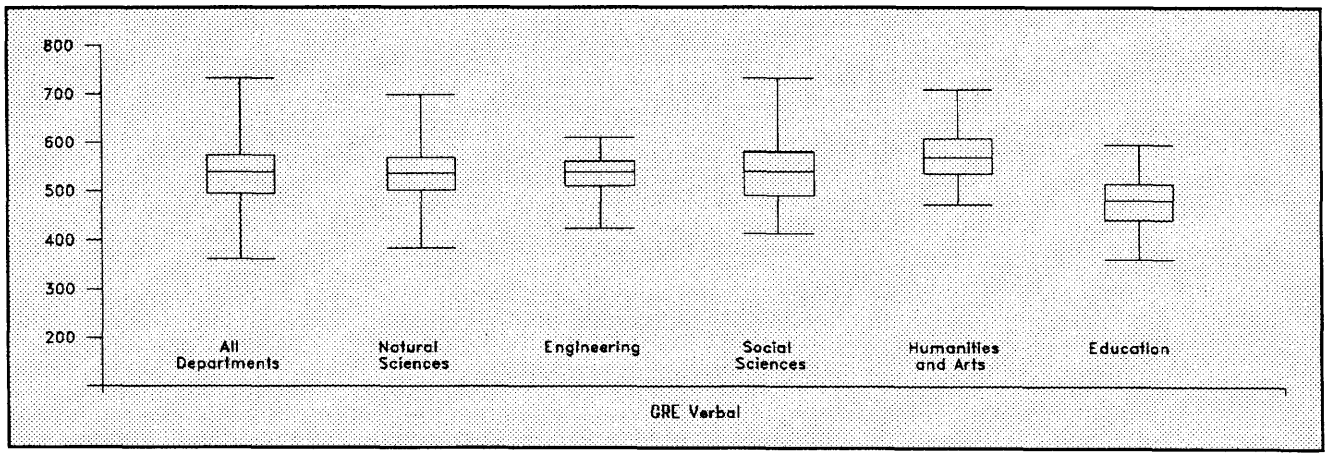



*Figure B.2*

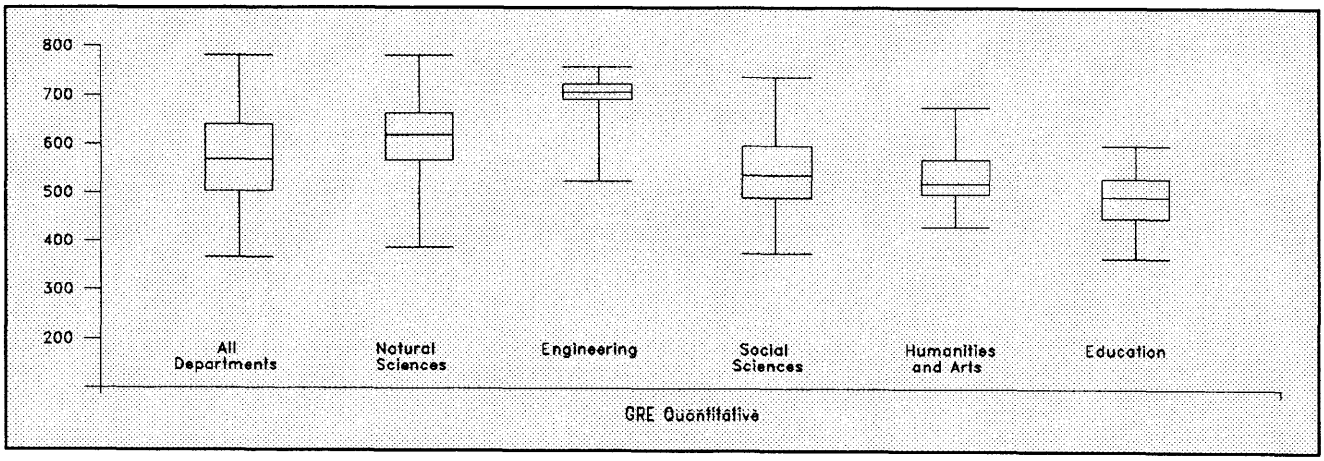
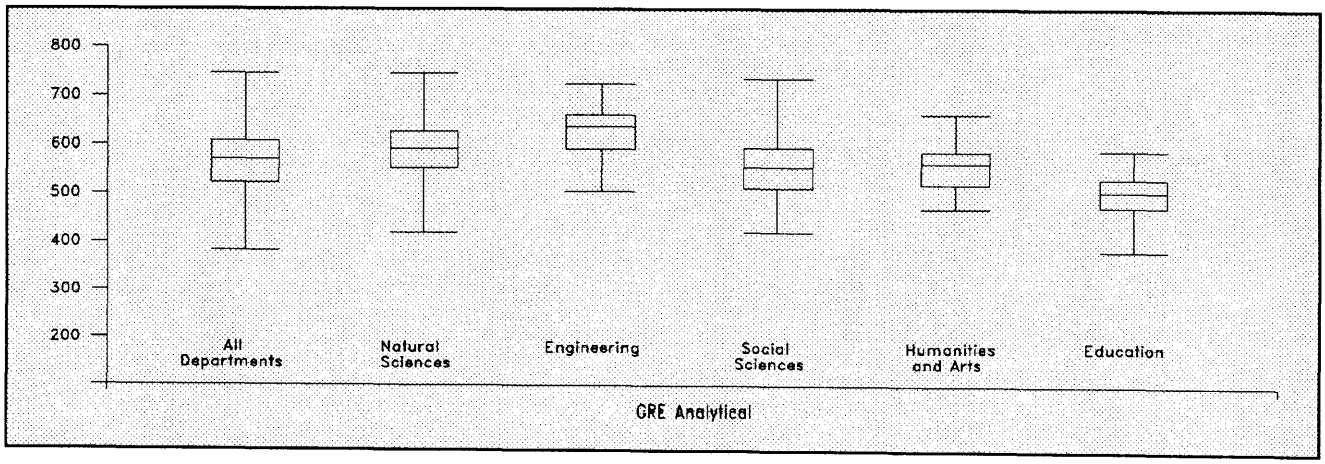



*Figure B.3*



# Figures B.1 - B.5

## Validity Sample : Students Who Communicate Better in English (Cont'd)

*Figure B.4*

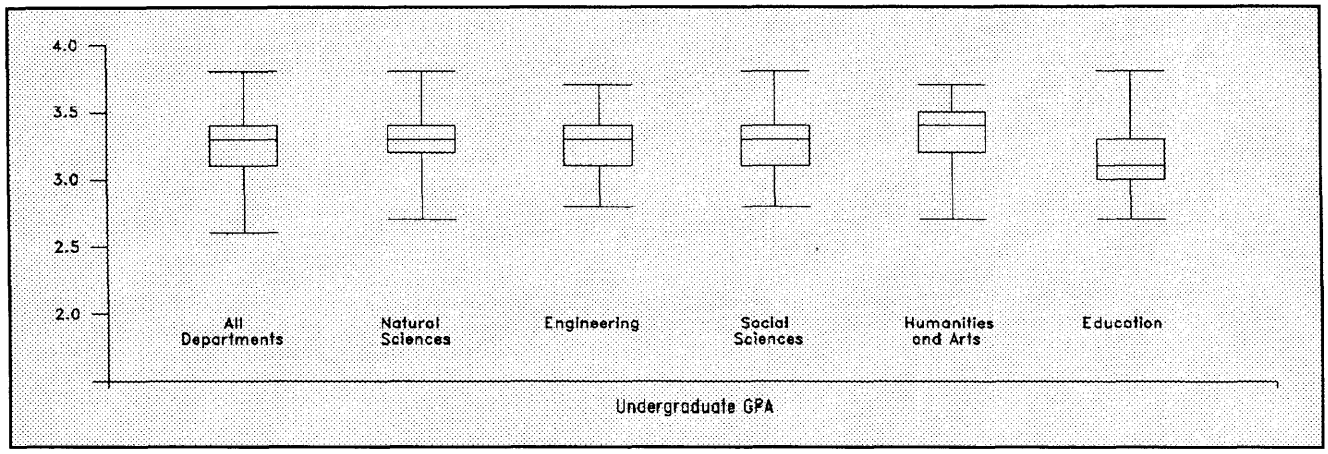
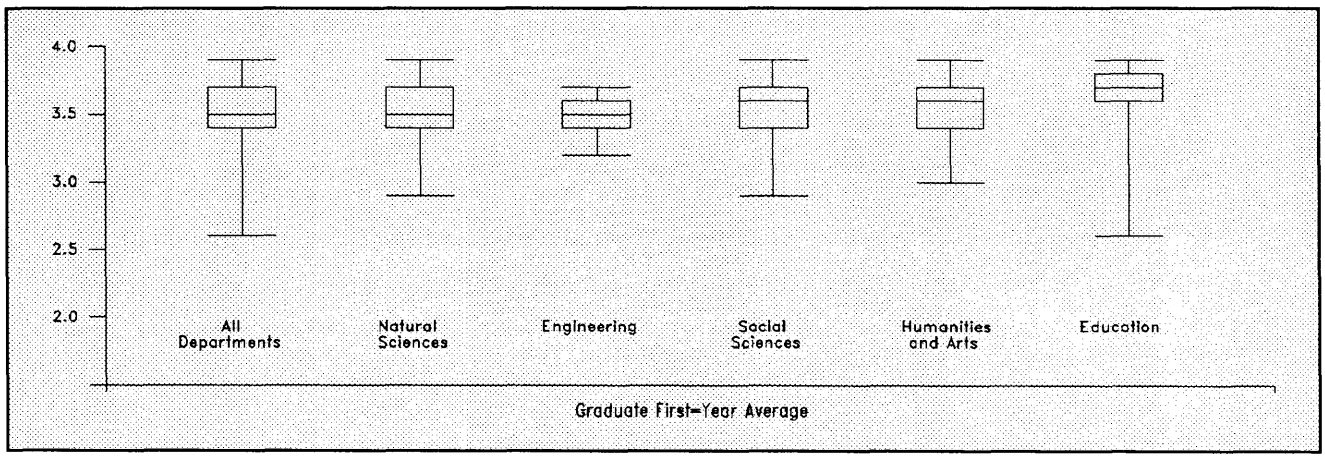



*Figure B.5*



# Figures B.6 - B.9

## Validity Sample : Students Who Do Not Communicate Better in English

*Figure B.6*

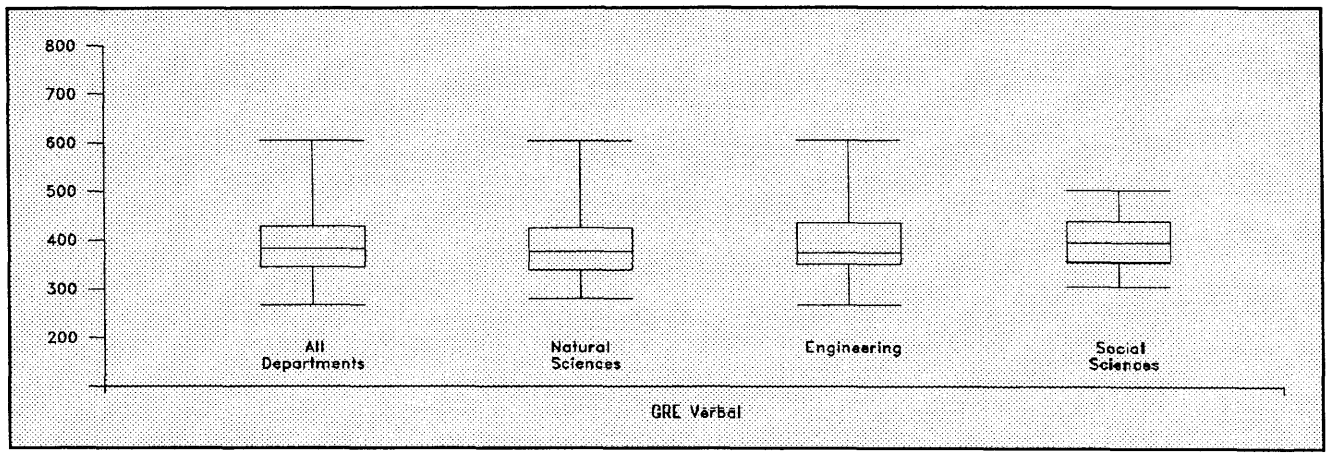
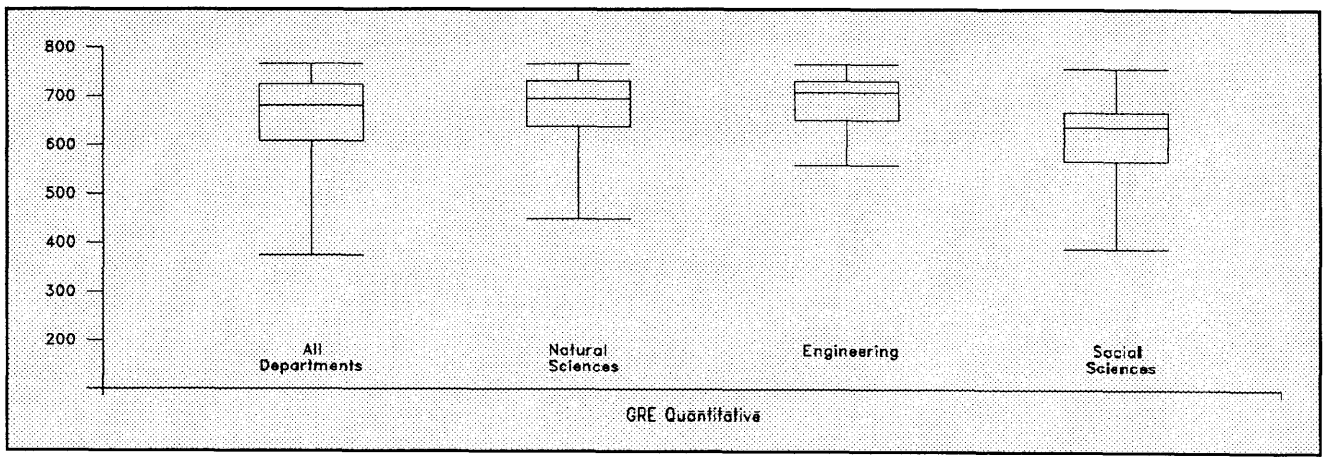



*Figure B.7*



# Figures B.6 - B.9

## Validity Sample : Students Who Do Not Communicate Better in English (Cont'd)

*Figure B.8*

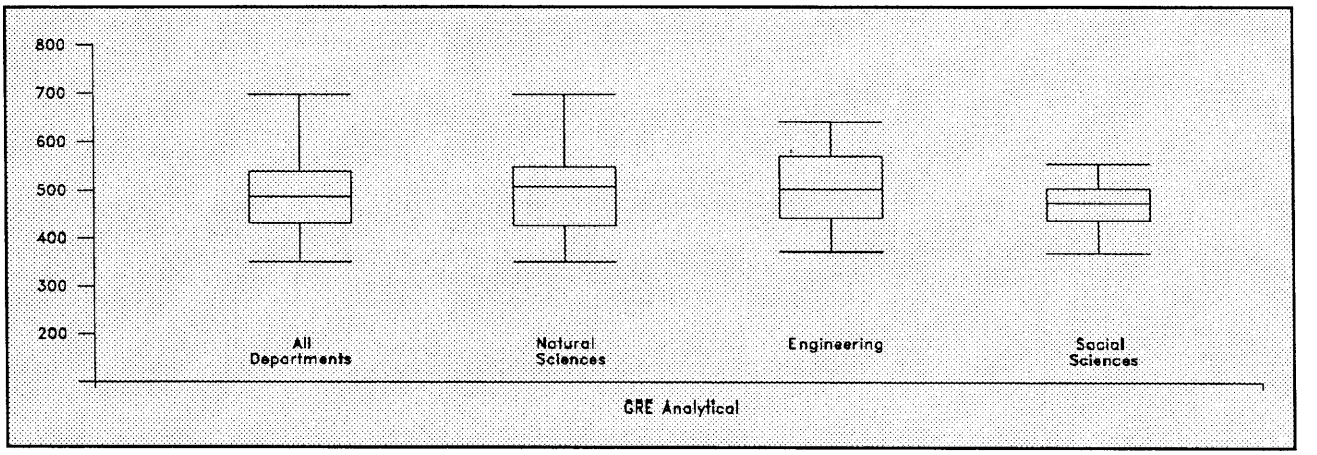
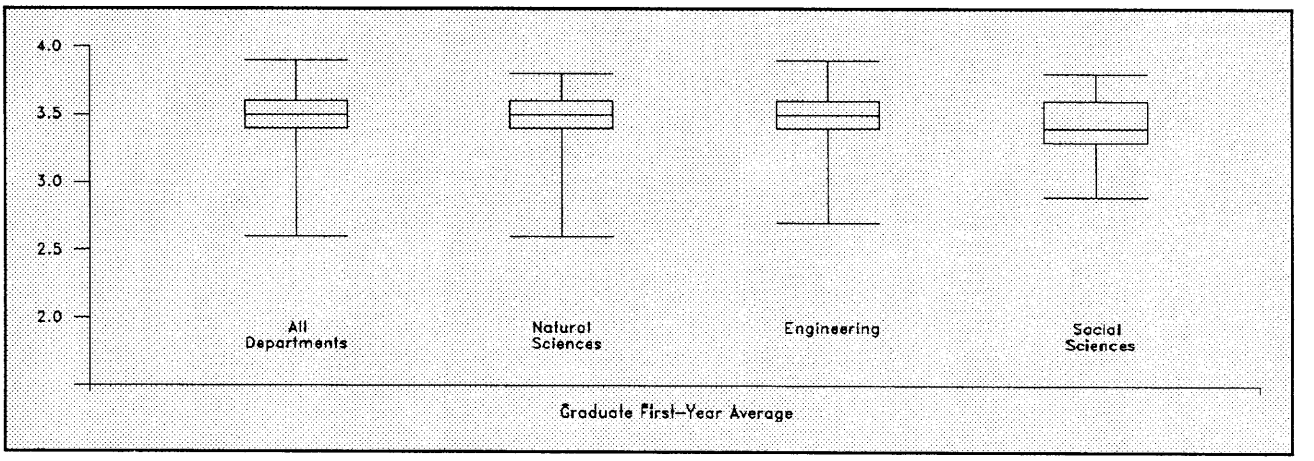



*Figure B.9*



Figures B.10 - B.11

Validity Sample : Students Who Communicate Better in English

*Figure B.10*

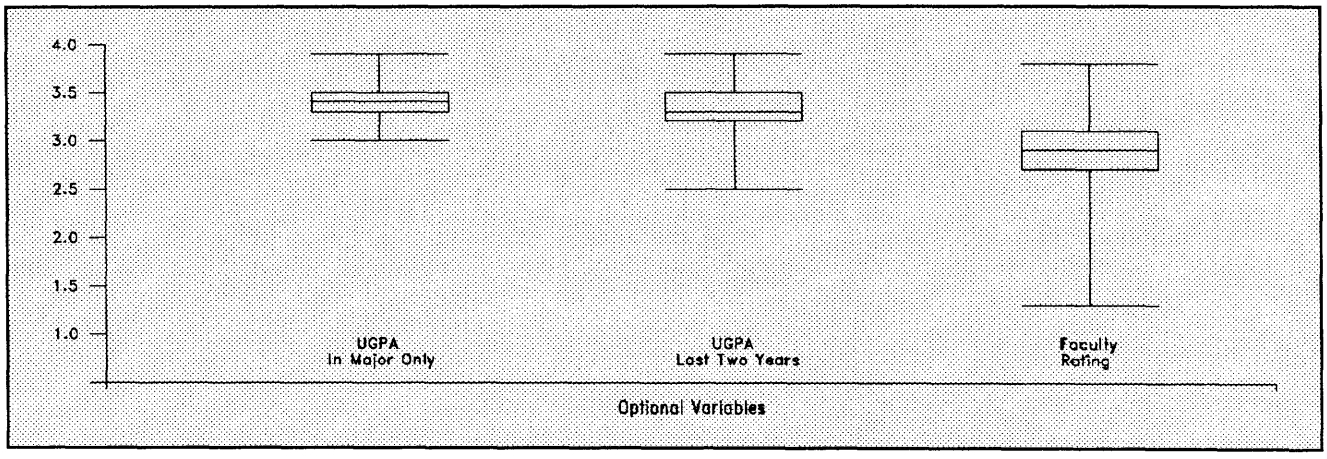
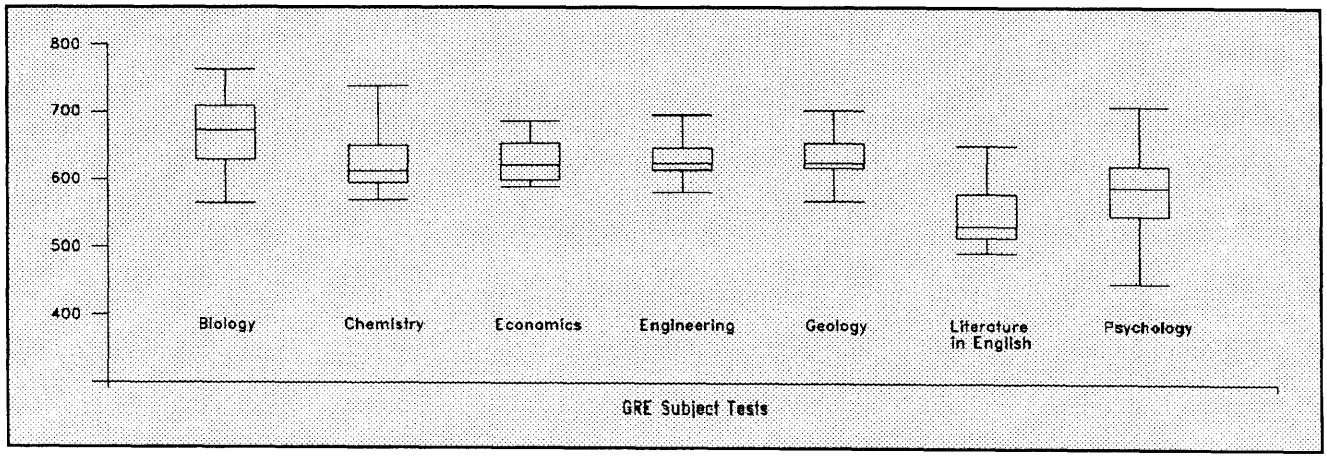



*Figure B.11**




* Physics data included in Table 7 are based on previous analyses and therefore are not included in this figure.

# Appendix C*

## Classification of Participants by Subgroups

Table C.1

Number of Students by Department Type and Sex
Validity Sample : Students Who Communicate Better in English

| | Males | | Females | | Total*** | |
|---|---|---|---|---|---|---|
| | Number | % | Number | % | Number | % |
| All Departments | 4,498 | 49 | 4,700 | 51 | 9,200 | 100 |
| Natural Sciences | 1,570 | 17 | 1,498 | 16 | 3,069 | 33 |
| Engineering | 359 | 4 | 83 | 1 | 442 | 5 |
| Social Sciences | 1,471 | 16 | 1,894 | 21 | 3,365 | 37 |
| Humanities & Arts | 724 | 8 | 541 | 6 | 1,266 | 14 |
| Education | 321 | 3 | 639 | 7 | 960 | 10 |
| Business | 53 | 1 | 45 | ** | 98 | 1 |

** Less than half a percentage point.
*** Sex information was not available for two students; therefore, the sum of a row may not equal the total listed.

Validity Sample : Students Who Do Not Communicate Better in English

| | Males | | Females | | Total** | |
|---|---|---|---|---|---|---|
| | Number | % | Number | % | Number | % |
| All Departments | 769 | 80 | 190 | 20 | 960 | 100 |
| Natural Sciences | 267 | 28 | 94 | 10 | 361 | 38 |
| Engineering | 306 | 32 | 22 | 2 | 329 | 34 |
| Social Sciences | 188 | 20 | 61 | 6 | 249 | 26 |
| Humanities & Arts | 8 | 1 | 13 | 1 | 21 | 2 |

** Sex information was not available for one student; therefore, the sum of a row may not equal the total listed.

* The percentages throughout this Appendix are based on the total number of departments or students. For the subgroup that communicates better in English, the totals for departments and students are 606 and 9,200, respectively. For the subgroup that does not communicate better in English, the totals for departments and students are 100 and 960, respectively.

Table C.2

Number of Students by Department Type and Age
Validity Sample : Students Who Communicate Better in English

| | <26 Years | | 26-30 Years | | 31-35 Years | | >35 Years | | Total** | |
|---|---|---|---|---|---|---|---|---|---|---|
| | No. | % | No. | % | No. | % | No. | % | No. | % |
| All Departments | 5,335 | 58 | 1,819 | 20 | 996 | 11 | 1,021 | 11 | 9,200 | 100 |
| Natural Sciences | 1,982 | 22 | 587 | 6 | 275 | 3 | 219 | 2 | 3,069 | 33 |
| Engineering | 346 | 4 | 66 | 1 | 23 | * | 6 | * | 442 | 5 |
| Social Sciences | 1,925 | 21 | 645 | 7 | 387 | 4 | 392 | 4 | 3,365 | 37 |
| Humanities & Arts | 669 | 7 | 286 | 3 | 150 | 2 | 159 | 2 | 1,266 | 14 |
| Education | 370 | 4 | 211 | 2 | 150 | 2 | 225 | 2 | 960 | 10 |
| Business | 43 | * | 24 | * | 11 | * | 20 | * | 98 | 1 |

* Less than half a percentage point.
** Age was not available for 29 students; therefore, the sum of a row may not equal the total listed.

Validity Sample : Students Who Do Not Communicate Better in English

| | <26 Years | | 26-30 Years | | 31-35 Years | | >35 Years | | Total** | |
|---|---|---|---|---|---|---|---|---|---|---|
| | No. | % | No. | % | No. | % | No. | % | No. | % |
| All Departments | 457 | 48 | 354 | 37 | 98 | 10 | 50 | 5 | 960 | 100 |
| Natural Sciences | 189 | 20 | 122 | 13 | 32 | 3 | 18 | 2 | 361 | 38 |
| Engineering | 167 | 17 | 125 | 13 | 26 | 3 | 11 | 1 | 329 | 34 |
| Social Sciences | 91 | 9 | 102 | 11 | 38 | 4 | 17 | 2 | 249 | 26 |
| Humanities & Arts | 10 | 1 | 5 | 1 | 2 | * | 4 | * | 21 | 2 |

* Less than half a percentage point.
** Age was not available for one student; therefore, the sum of a row may not equal the total listed.

Table C.3

Number of Departments by Department Type and Department Size
Validity Sample : Students Who Communicate Better in English

| | 5-9 Students | | 10-14 Students | | 15-19 Students | | >19 Students | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|
| | No. | % | No. | % | No. | % | No. | % | No. | % |
| All Departments | 229 | 38 | 154 | 25 | 87 | 14 | 136 | 22 | 606 | 100 |
| Natural Sciences | 98 | 16 | 63 | 10 | 25 | 4 | 43 | 7 | 229 | 38 |
| Engineering | 12 | 2 | 12 | 2 | 5 | 1 | 5 | 1 | 34 | 6 |
| Social Sciences | 73 | 12 | 42 | 7 | 34 | 6 | 54 | 9 | 203 | 34 |
| Humanities & Arts | 26 | 4 | 17 | 3 | 14 | 2 | 17 | 3 | 74 | 12 |
| Education | 18 | 3 | 16 | 3 | 8 | 1 | 16 | 3 | 58 | 10 |
| Business | 2 | * | 4 | 1 | 1 | * | 1 | * | 8 | 1 |

* Less than half a percentage point.

Validity Sample : Students Who Do Not Communicate Better in English

| | 5-9 Students | 10-14 Students | 15-19 Students | >19 Students | Total |
|---|---|---|---|---|---|
| | No. | No. | No. | No. | No. |
| All Departments | 60 | 26 | 9 | 5 | 100 |
| Natural Sciences | 27 | 8 | 4 | 1 | 40 |
| Engineering | 15 | 9 | 4 | 3 | 31 |
| Social Sciences | 15 | 9 | 1 | 1 | 26 |
| Humanities & Arts | 3 | 0 | 0 | 0 | 3 |

* Because the total number of departments in the subgroup that does not communicate better in English is 100, the percentage is the same as the number of departments.

Table C.4

Number of Students by Department Type and Ethnic Group
Validity Sample : Students Who Communicate Better in English

| | American Indian Alaskan Native | | Asian/ Pacific Islander | | Black | | Hispanic | | White | | Other | | Total** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | No. | % | No. | % | No. | % | No. | % | No. | % | No. | % | No. | % |
| All Departments | 76 | 1 | 171 | 2 | 313 | 3 | 164 | 2 | 8,356 | 91 | 78 | 1 | 9,200 | 100 |
| Natural Sciences | 53 | 1 | 75 | 1 | 60 | 1 | 49 | 1 | 2,789 | 30 | 21 | * | 3,069 | 33 |
| Engineering | 2 | * | 19 | * | 10 | * | 13 | * | 393 | 4 | 5 | * | 442 | 5 |
| Social Sciences | 16 | * | 60 | 1 | 180 | 2 | 68 | 1 | 2,995 | 33 | 30 | * | 3,365 | 37 |
| Humanities & Arts | 1 | * | 12 | * | 41 | * | 24 | * | 1,179 | 13 | 9 | * | 1,266 | 14 |
| Education | 3 | * | 3 | * | 19 | * | 5 | * | 916 | 10 | 10 | * | 960 | 10 |
| Business | 1 | * | 2 | * | 3 | * | 5 | * | 84 | 1 | 3 | * | 98 | 1 |

* Less than half a percentage point.
** Ethnicity information was not available for 42 students; therefore, the sum of a row may not equal the total listed.

Validity Sample : Students Who Do Not Communicate Better in English

| | American Indian Alaskan Native | | Asian/ Pacific Islander | | Black | | Hispanic | | White | | Other | | Total** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | No. | % | No. | % | No. | % | No. | % | No. | % | No. | % | No. | % |
| All Departments | 2 | * | 500 | 52 | 20 | 2 | 111 | 12 | 71 | 7 | 246 | 26 | 960 | 100 |
| Natural Sciences | 2 | * | 200 | 21 | 5 | 1 | 49 | 5 | 25 | 3 | 80 | 8 | 361 | 38 |
| Engineering | 0 | 0 | 159 | 17 | 6 | 1 | 28 | 3 | 25 | 3 | 111 | 12 | 329 | 34 |
| Social Sciences | 0 | 0 | 138 | 14 | 9 | 1 | 26 | 3 | 19 | 2 | 47 | 5 | 249 | 26 |
| Humanities & Arts | 0 | 0 | 3 | * | 0 | 0 | 8 | 1 | 2 | * | 8 | 1 | 21 | 2 |

* Less than half a percentage point.
** Ethnicity information was not available for ten students; therefore, the sum of a row may not equal the total listed.

Table C.5*

Number of Departments with One or More Students in Ethnic Group

| | Communicates Better in English | | Does Not Communicate Better in English |
|---|---|---|---|
| | Number | % | Number** |
| American Indian/Alaskan Native | 30 | 5 | 1 |
| Asian/Pacific Islander | 122 | 20 | 75 |
| Black | 177 | 29 | 17 |
| Hispanic | 110 | 18 | 41 |
| White | 598 | 99 | 44 |
| Other | 59 | 10 | 55 |

* This table shows the number and percentage of departments that have at least one student in an ethnic group. For example, 177 departments in the subgroup that communicates better in English have at least one Black student. Departments may be included more than once.

** Because the total number of departments in this group is 100, the percentage is the same as the number of departments listed for each ethnic group.